



A Survey Note on Big Data Analytics Frameworks on the Hadoop Platform

Y. Tayar¹, S. Satyanarayana² and R. Sivaram Prasad³

Department of Computer Science and Engineering

Lingayas' Institute of Management and Technology, Madalavarigudem, Vijayawada Rural,
Andhra Pradesh, India 521212

1. Corresponding Author, Email: tayar.lingayas@limat.edu.in, tayarjan6@gmail.com
2. Department of AI & ML, Mallareddy University, Hyderabad, Telangana State 500100, E-Mail: drsatanarayana@mallareddyuniversity.ac.in
3. Department of Commerce and Management Studies, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India 522508. Email: raminenisivaram@yahoo.co.in

Abstract

Data analysis has become one of the primary tools in big data business companies for solving problems and Hadoop has been the preferred platform to achieve this task. In different industrial as well as public sectors like finance management, health care sector and agro-industries, Hadoop has potentially changed the face of data analytics for providing the better services as also resolved various customer and management oriented complications. A survey of the secondary data was adopted in the present study wherein the participants were asked about their opinion in a questionnaire format. The study revealed that Hadoop is versatile and cost-effective in storing and processing of structured as well as unstructured data. Computational cost of a clustering algorithm could be significantly lowered by the use of fundamental mathematical ideas, and the computation time for a data analysis method would decrease through the concurrently using distributed computing. In a way Hadoop has proved successful in data processing and commercial operations and continues to be the preferred platform due to its abilities of using distributed file system to manage data storage, reading the huge data from one machine in big data analytics despite few potential limitations.

Key words: Big Data Analytics, Hadoop Platform, Distributed file system, Open-source framework

DOI: 10.48047/ecb/2023.12.si12.121

Introduction

Despite the development of computing hardware as a result of advances in computer systems and internet technologies, limitations in handling large-scale data management still persist even

in the big data era since most of the conventional data analytics could not be subjected for direct analysis of huge data (Fisher *et. al.*, 2012; Tsai *et. al.*, 2015). Out of various data analytics and frameworks that have been framed recently, Hadoop has become popular in the modern digital era in solving problems in data analysis in big data business companies (Husain and Zeebaree, 2021). In different industrial and public sectors, finance management, health care sector (Dash *et. al.*, 2019) and agro-industries (Wolfret *et. al.*, 2017), Hadoop and Hadoop relevant technologies have changed the face of data analytics in finding better solution than other methods (Reddy and Kumar, 2009). Despite various limitations from the early days Hadoop has been always successful in managing the data business (Sandhya and Xiao, 2019) and processing tasks (Fakherldin *et al.*, 2019). Some vulnerabilities arise due to the flexibility of framework and lead to threats to the data and lead to attacks (Singh and Singh, 2019). This study has been framed to study overall preference of Hadoop popularity in Big Data Analytics in designing solutions using a survey method with objectives of (i) understanding the working applications of the Hadoop platform in data analytics, (ii) evaluating the acceptance and relevance of Hadoop in the digital era, (iii) determining the advantages of the Hadoop platform in huge data analytics and understanding its limitations and shortcomings in big data analysis.

Methodology

The research methodology in this study contained Hadoop Big Data Analytics Market research carried out using secondary data to collect information about the Hadoop platform and various data analytics companies' work performance (Jin, 2016). The survey method was chosen with the objective of establishing suitability of Hadoop platform in data analytics by testing through the following hypotheses.

1. Hadoop platform helps in the establishment of data analysis operations.
2. Open-source framework-based software performs efficiently than other systems.
3. Hadoop's advantages far outweigh the problems self-generated by Hadoop

Seventy Five participants were asked about their opinion in a questionnaire format comprising narrative questions and the responses were analyzed using Durbin Watson Statistic and the results are tabulated as also presented in figure format wherever necessary (Anderson, 2000; Sivarajah *et. al.*, 2016).

Results and Discussion

Hadoop has gained significant market increase in the past five years and it records the changes in the size of the market from the year 2017 to 2023 involving monetary escalation from 17 to 99.3 million USD with steep increase of 500% (Fig. 1). The global Hadoop market is

anticipated to reach USD 884.35 billion by 2030 (approx. INR 725,000 crore) due to the widespread use of Hadoop-dependent applications for real-time analytics and web-based business practices. The market for Hadoop is expanding in China, India, Japan, South Korea, Australia and the rest of Asia-Pacific due to increased internet penetration and technology advancements as also digital infrastructure (Anonymous, 2022). Current study discusses the effectiveness of the Hadoop platform in big data analytics and the market. Solving various problems in data analysis is the major concern for big data business companies and Hadoop has been their preferred platform to achieve this task (Jenhani *et al.*, 2019). In different industrial sectors and public sectors like finance management, health care sector, and agro industries Hadoop has changed the face of data analytics for the better. From the data collected from the secondary collection method it was found that the Hadoop platform helps in establishment of data analysis operations and open source framework based software are always better than others (Suryanarayana *et al.*, 2019; Wilson, 2020).

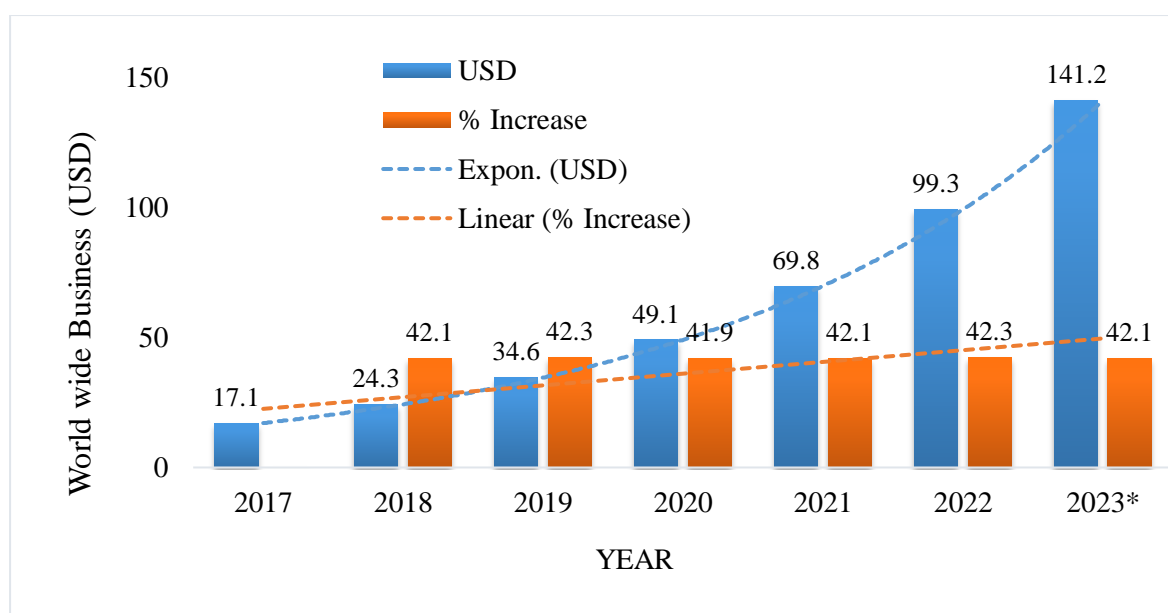


Fig. 1. Worldwide business of Hadoop and Big Data and percent of Increase over previous years during 2017-2023 (2017= treated as base year, 2023*= projected value).

Demographic Analysis: Gender

The present study included 75 respondents in which there were 65.3% male (49/75) and 33.4% female (25/75). There was one respondent (1.3%) who belonged to transgender community (Fig. 2). But, the respondent has requested for anonymity because of stigma of facing social discrimination (Philip and Soumyaja, 2019) despite the fact that various courts have admitted the transgender people as a 'third gender' (Mittal and Garg, 2015). Transgender community is

slowly finding way into mainstream employment because several companies have started hiring Transgender people with an idea of introducing diversification in work place (Sawant, 2017; Magenes, 2021). Gender diversity issue seems to be critical in software companies wherein women Technologists accounted for 27.6% (James *et. al.*, 2017). This situation is more fragile in big companies where female Software Engineers were accounted for 22% as against 25% in small companies in countries like United States (Gregory *et. al.*, 2023) while women workers in the IT firms in India account to 34% (Anonymous, 2017). Canedo *et. al.* (2017) found that the situation is mainly because of (i) lack of interest among women, (ii) gender bias and (iii) lack of support for women. Thus, participation of women in software remains unbalanced despite various efforts to improve diversity (Trinkenreich *et. al.*, 2019). But, situation in India is encouraging that women in software industry make up from 34.2 to 36.3% in some companies (Raja and Sivakumar, 2022). Chakraborty (2019) reported that although Indian software industry comprise 35% women employees, most of them are stagnated in entry level jobs and as a result, availability of experienced female candidates is low. Accordingly, Kumar (2023) found that the proportion of women working as heads in Information Technology is dropping gradually. Further, in countries like India, Gender inequality extends beyond economic growth and manifests in the form of socially controlled and preimposed gender roles in the society (Batra and Reio, 2016).

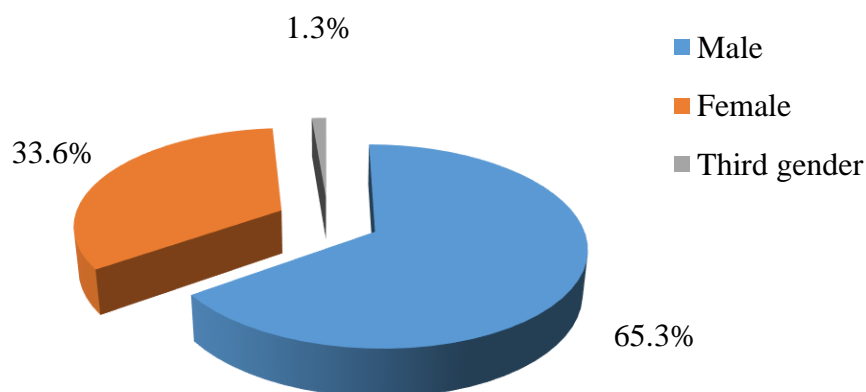


Fig. 2. Pie chart showing sex of respondents who participated in the survey on “Big Data Analytics Frameworks on the Hadoop Platform”

Demographic Analysis: Age Group

Participants in the present study belonged to different age groups ranging from 25 to 60 years in which maximum employees (40/75) represented the age group of 46 to 60 years (60%) followed by 26.7 % (20/75) who fell in the age group of 36 to 45 while only 13.3% (10/75)

belonged to the age group of 25 to 35 (Fig. 3). Sankar and Changat (2017) found that the average age of those working in the Indian IT industry is 42.5 years. Comparable situation prevailed in Europe wherein 48.2% of software engineers in Europe fall in the age group ranging from 25 to 34 (Dimitrievski, 2023). This is an alarming situation that ageing would be one of the strongest reasons that results in declining working age population as well as productivity in Europe by 4% due to ageing (Smitet. al., 2020). Study conducted by Scalers (2022) in The United States also revealed that average age of IT professionals in The United States is 41.3 years. In a study conducted elsewhere, Nasehi (2013) found that 30.9% of the employees are in the age group of 35-44 and those who are in the age group 50+ are only 17.60% and opined that quantum of ageing employees in IT industry will significantly affect the production process. However, during the pandemic, digital technologies have started to play unprecedentedly prominent role with increase in the use of digital platforms and thus resulted in retaining the employees albeit ageing (Cepal, 2021).

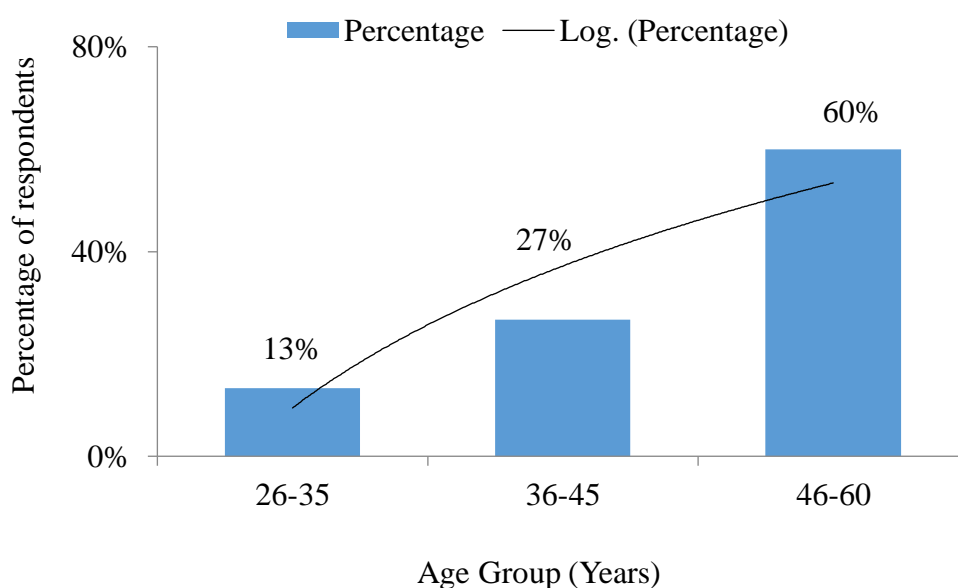


Fig. 3. Histogram with log-curve showing age group of respondents who participated in the survey on “Big Data Analytics Frameworks on the Hadoop Platform”

Demographic Analysis: Income levels

Annual income levels of participants in the current study ranged from ₹ 2.40 to 7.20+ lakh. Cross section analysis of respondents revealed that 66.7% of them (50 out of 75) fall within the annual income range of ₹ 2.40 to 4.00 lakh followed by 20% (15 out of 75) of respondents

whose annual income ranged from ₹ 49.20 to 7.20 lakh and remaining 13.3% respondents were drawing annual income above ₹ 7.20 lakh (Fig. 4). Mohamed Saeed *et. al.* (2023) reported that skills such as language and quantitative aptitude, as well as attitude towards complete a task contribute significantly to the starting pay of engineering graduates in India while Verma (2023) reported that software engineering is a highly preferred career choice due to high demand and rewarding career profiles with an average salary of ₹ 3 to 14 LPA in India.

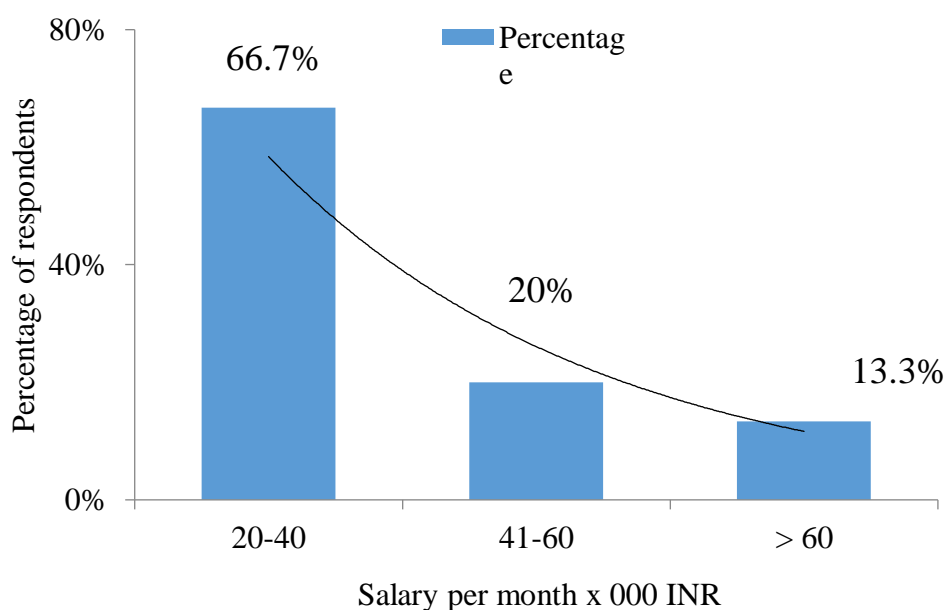


Fig. 4. Histogram showing income levels of respondents who participated in the survey on Big Data Analytics Frameworks on the Hadoop Platform.

Hadoop platform and Data Analysis

Durbin Watson value obtained for the assumption that “Hadoop platform helps in the establishment of data analysis operations” is 1.75 with a significant F change value of 0.000. The value of R and R square is 0.677 and 0.459 respectively. ANOVA analysis revealed F value of 20.080 supported by significant value of 0.000 thus establishing positive correlation between Hadoop platform and data analysis operations (Table 1). Sethy and Panda (2015) recorded that Hadoop processing involves various important tasks in which data is divided into directories and files, subsequently into blocks before putting through it for final processing. This makes Hadoop into a unique big data analysis platform when compared to other systems. Mazumder and Dhar, (2016) have stated that although, the Google File System (GFS) and the Hadoop Distributed File System (HDFS) are both based on the GFS, HDFS distinguishes itself by

displaying high throughput access to analytic data containing huge datasets (Merceedi and Sabry, 2021).

Table-1. ANOVA analysis and Durbin Watson Statistic of data accrued “Hadoop platform helps in the establishment of data analysis operations”

Model Summary ^b										
Model	R	R ²	Adjusted R ²	SE	Change Statistics					Durbin Watson Statistic
					R ² Change	F Change	df1	df 2	Significant F Change	
1	.065 ^a	.004	.038	.907	.004	.101	3	71	.959	2.216

ANOVA ^a						
Model	Sum of Squares	Sum of Squares	dF	Mean Square	F	Significance level
1	Regression	.250	3	.083	.101	.959 ^b
	Residual	58.416	71	.823		
	Total	58.667	74			

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Significance level
		B	Standard Error	Beta		
1	Constant	3.680	.538		6.836	.000
	1.1	.041	.147	.050	.277	.782
	1.2	.033	.128	.047	.259	.796
	1.3	.045	.099	.055	.459	.648

Hadoop platform is Open-source framework

Similarly, Durbin Watson Statistic for the hypothesis that “Open-source framework-based software is better than others” was calculated to be 1.75 with a significant F change value of 0.000 while value of R and R square was found to be 0.677 and 0.459 respectively. ANOVA analysis yield an F value of 20.080 following significant value of 0.000 and confers positivity that open-source framework-based software yields better results than others (Tale 2). Corbly (2014) suggested that an open-source software contain various criteria such as software is cost free, the program is distributed in user friendly form so that programmers and/or end-users will be able to use it similar to the original version of the software. In contrast to proprietary software or closed source software, the adoption of frameworks based on Free Open-Source Software (FOSS) like Hadoop has emerged as a competitive alternative because it fully meets the end-users' needs without requiring any adaptation and modifications (Martinez-Torres and Diaz-Fernandez, 2014; Teixeira *et. al.*, 2015).

Table-2. ANOVA analysis and Durbin Watson Statistic of data accrued “Open-source framework-based software performs efficiently than other systems”

Model Summary ^b										
Model	R	R ²	Adjusted R ²	SE	Change Statistics					Durbin Watson Statistic
					R ² Change	F Change	df1	df2	Significance F Change	
1	.677 ^a	.459	.436	.713	.459	20.08	3	71	.000	1.755

Model Summary ^b										
Model	R	R ²	Adjusted R ²	SE	Change Statistics					Durbin Watson Statistic
					R ² modified	F Change	df1	df2	Significance F Change	
1	.677 ^a	.459	.436	.713	.459	20.08	3	71	.000	1.755

Model Summary ^b										
Model	R	R ²	Adjusted R ²	SE	Change Statistics					Durbin Watson Statistic
					R ² modified	F Change	df1	df2	Significance F Change	
1	.179	.032	.005	.888	.032	1.197	2	72	.308	2.209

Hadoop is critical

Regression analysis of “Hadoop’s advantages far outweigh the limitations that are innately possessed by Hadoop” yields Durbin Watson value of 1.75 with a significant F change value of 0.000 was obtained while The values of R and R square are 0.677 ad 0.459 respectively. ANOVA analysis resulted a F value of 20.080 and a significant value of 0.000 which substantiate that the Hadoop’s advantages minimizes the shortfalls that exist within Hadoop (Table 3). The core framework and structure of Hadoop are mainly formed by components which are known as file systems in distributed form, Map reducing structure of Hadoop and the YARN. Hadoop is a platform which can work without using a super computer with heavy processing power and instead, it can work with scalability that is so impressive and can use one computer to hundreds and thousands of computers at once (Omar and Juna, 2019). This feature makes Hadoop very much preferable in current times where alternative platforms have also come into the market (Yadav and Chandra, 2017). Hadoop stores and process the voluminous data using many modules and the resources, configures and retrieves the data using highly fault tolerance mechanism (Suguna and Devi, 2016). Kaur *et. al.* (2016) records that the Hadoop Distributed File System (HDFS) store massive amount of data and keep working even when significant elements of the storage infrastructure fail. Big data or information described in an informative manner is related to big data and also a binding impact force among the different

sectors of advancement in social and economic forums (Yao, 2020). Fast-developing data analysis and information theories assist organizations in storing data about manufacturing, customer management, finance solutions and tracking customer activities. The Hadoop platform has always fallen short in cases where files are generally small and surrounding the data in a small capacity (Priyanka *et al.*, 2020). The Hadoop platform always needs processing power which is high and could only handle batch processing tasks and the security issues of Hadoop have also been a critical concern (Bathal and Singh, 2019).

Table-3. ANOVA analysis and Durbin Watson Statistic of data accrued “Hadoop’s advantages far outweigh the problems self-generated by Hadoop”

Model Summary ^b											
Model	R	R ²	Adjusted R ²	SE	Change Statistics					Durbin Watson Statistic	
					R ² Change	F Change	df1	df2	Significance F Change		
1	.179 ^a	.032	.005	.888	.032	1.197	2	72	.308	2.209	

ANOVA ^a							
Model	Sum of Squares	Sum of Squares	dF	Mean Square	F	Significance level	
1	Regression		1.888	2	.944	1.197	.308 ^b
	Residual		56.779	72	.789		
	Total		58.667	74			

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Significance level
		B	Standard Error	Beta		
1	Constant	3.107	.505		6.158	.000
	3.1	.059	.096	0.76	.611	.563
	3.2	.141	.141	.136	1.092	.278

Conclusion

Hadoop framework plays a significant role in storing and processing of Big Data besides providing fast and cost-effective solution. It finds wide application in sectors like telecommunication, healthcare and government sectors, insurance and social media. Although Hadoop there are few limitation in the Hadoop platform like small files, vulnerability and high end processing but still Hadoop is still the preferred platform for many big data companies. Different research studies on Hadoop imply that it can significantly influence security scaling in a way that keeps up with the volume of data that it protects. As a result, it's necessary to update the Hadoop system with security measures without installing and configuring it

separately. Barring a few security issues, critical and accurate throughput of Hadoop is very advantageous for data analytics companies wherein the Big Data platform will achieve the biggest rise in adoption.

Acknowledgments

Authors (YT) gratefully acknowledges the support received from the authorities of Lingayas' Institute of Management and Technology in conducting the current work. Authors also acknowledge the significance of the publications that have been referred and quoted in this article.

References

- Anonymous, 2017. Tech companies have only 26% women in engineering roles. *Economic Times*, 10:22. https://economictimes.indiatimes.com/tech/ites/tech-companies-have-only-26-women-in-engineering-roles.Survey/articleshow/61171280.cms?utm_source=content_of_interest&utm_medium=text&utm_
- Anonymous, 2022. Information by Component (Hardware, Software), Deployment Model (On-Premise, Cloud, and Hybrid) and Region-Forecast till 2030. <https://straitresearch.com/report/hadoop-market>.
- Bathal, G. S. and Singh, A. 2019. Big Data: Hadoop framework vulnerabilities, security issues and attacks. *Array*, 1-2: 8.
- Batra, R. and Reio, T. G. 2016. Gender Inequality Issues in India. *Advances in Developing Human Resources*, **18(1)**: 88–101.
- Canedo, E. D., Bogo, M., Fagundes, F. and José Antonio. S. C., 2017. Barriers Faced by Women in Software Development Projects. *Information (Switzerland)*, 309-329.
- Cepal, N. U. 2021. Digital Technologies for new future. United Nations publication, ELAC, 95pp.
- Chakraborty, S. 2019. The business case for gender diversity in the Indian information Technology industry. In *Inequality and Organizational Practice* (pp. 211-233). Palgrave Macmillan, Cham.
- Corbly J. E. 2014. The Free Software Alternative: Freeware. *Open-source Software and Libraries*. 2014, 65–76.
- Dash, S., Shakyawar, S. K., Sharma, M. and Kaushik, M. 2019 Big data in healthcare: management, analysis and future prospects. *Journal of Big Data* **6**: 54. <https://doi.org/10.1186/s40537-019-0217>.
- Dimitrievski, M. 2023. Software Development Statistics – 2023. <https://truelist.co/blog/software-development-statistics>.
- Fakherldin, M. A. I., Adam, K., Bakar, N. A. A., and Majid, M. A. 2019. Weather data analysis using Hadoop: applications and challenges. In *IOP Conference Series: Materials Science and Engineering* (Vol. 551, No. 1, p. 012044). IOP Publishing.

- Fisher D, DeLine R, Czerwinski M, Drucker S. 2012. Interactions with big data analytics. *Journal of Interactions*, 19(3): 50–9.
- Gregory, M., Nelson, C. and Dixon, L. 2022. Progress for Gender equality in the Tech Industry. *Percentage of Female Software Engineers. Celential.ai*, 5:8.
- Husain, B. H. and Zeebaree, S. R. 2021. Improvised distributions framework of hadoop: A review. *International Journal of Science and Business*, 5(2), 31-41. <https://ideas.repec.org/aif/journal/v5y2021i2p31-41.html>
- Hussain, T., Sanga, A. and Mongia, S. 2019. Big Data Hadoop Tools and Technologies: A Review. *Computer Sc.*,
- James, T., Galster, M, Blincoe, K. and Miller, G. 2017. *What Is the Perception of Female and Male Software Professionals on Performance, Team Dynamics and Job Satisfaction? Insights from the Trenches. 13-22. 10.1109/ICSE-SEIP.2017.31.*
- Jenhani, F., Gouider, M. S., and Said, L. B. 2019. Streaming social media data analysis for events extraction and warehousing using hadoop and storm: Drug abuse case study. *Procedia Computer Science*, 159: 1459-1467.
- Jin, G. W. 2016. A Study on the Data Collection Methods based Hadoop Distributed Environment. *Journal of the Korea Convergence Society*, 7: 1-6.
- Kaur, I., Kaur, N., Ummat, A., Kaur, J. and Kaur. N. 2016. Research Paper on Big Data and Hadoop. *International J. Comp. Sc. and Technology*, **7(4)**: 50-53.
- Kumar, N. 2023. Gender Gap in Indian Technology sector. *NAASCOM Diversity and Inclusion*, **16**: 3.
- Magenes, F. 2021. Ten Companies that hired Transgender in India. <https://www.zerokaata.com/zerokaata-studio/hiring-transgenders-in-india/#indian-companies-that-hired-transgenders>
- Martinez-Torres, M. R., Diaz-Fernandez, M. C. 2013. Current issues and research trends on open-source software communities. *Technol. Anal. Strategy Manag.* **26**: 55–68.
- Mazumder, S. and Dhar, S. 2015. Hadoop as Big Data Operating System – The Emerging Approach for Managing Challenges of Enterprise Big Data Platform.
- Mercedi, K. J. and Sabry, N. A. 2021. A Comprehensive Survey for Hadoop Distributed File System. *Asian Journal of Research in Computer Science*, **11(2)**: 46-57.
- Mittal, K. and Garg, S. 2015. Transgender in India: New developments and enactments. *International Journal of Multidisciplinary Res. and Dev.*, **2(11)**: 402-405.
- Mohamed Saeed, A. K., Abdullah, P. and Tahir, A. T. 2023. Salary Prediction for Computer Engineering Positions in India. *J. Applied Science and Technology Trends*, **4(1)**: 13-18.
- Nasehi, A. 2013. A Quantitative Study on Critical Success Factors in Agile Software Development Projects; Case Study IT Company. Master's Thesis, University of Boras, 109pp.

- Omar, H. K. and Juma, A. K. 2019. Big data analysis using apache spark MLlib and Hadoop HDFS with Scala and Java. *Kurdistan Journal of Applied Research*, 4(1):7-14. <https://www.kjar.spu.edu.iq/index.php/kjar/article/view/265>.
- Peterson, R. A. 2000. Constructing Effective Questionnaires. SAGE Publications, Inc. 158pp.
- Philip, J. and Soumyaja, D. 2019. Work place diversity and Inclusions: policies and practices for organizations employing Transgender people in India. *Int. J. Public Policy*, 15(3/4): 299-313.
- Priyanka, E. B., Thangavel, S. and Prabu, D. V. 2020. Fundamentals of Wireless Sensor Networks Using Machine Learning Approaches: Advancement in Big Data Analysis Using Hadoop for Oil Pipeline System with Scheduling Algorithm. In: *Deep Learning Strategies for Security Enhancement in Wireless Sensor Networks* (pp. 233-254).
- Raja, S. and Sivakumar, I. 2022. Discovering Gender Bias in Indian IT Industry. 2:22-26. 10.55529/jwes.22.22.26. *Journal of Women Empowerment and Studies*, 2(2): 22-26.
- Reddy, D. S. S. and Kumar, U. R. 2019. Hadoop Framework for Big Data Analytics to Analyze Attacks on Business Data. *International Journal of Innovative Technology and Exploring Engineering*, **9(2)**: 3995-4000.
- Sandhya, A. and Xiao M. L. 2019. Big Data Analytics and Deep Learning in Bioinformatics With Hadoop. Deep Learning and Parallel Computing Environment for Bioengineering Systems. <https://www.sciencedirect.com/topics/computer-science/hadoop-platform>.
- Sankar, S. S. and Changat, S. 2017: Two Scenarios for the Indian IT Industry. *Journal of Futures Studies*, **22(1)**: 39–56.
- Sawant, N. 2017. Transgender: Status in India. *Annals of Indian Psychiatry*, **10**: 43-47.
- Scalers, 2022. Development deep dive: how many software engineers in the US?
- Sethy, R. and Panda, M. 2015. Big Data Analysis using Hadoop: A Survey. *International Journal of Advance Research in Computer Science and Software Engineering*. 5.
- Singh, G. B. and Singh, A. 2019. Big Data: Hadoop framework vulnerabilities, security issues and attacks. *Array*, 1&2: 100002.
- Sivarajah, U., Kamal, M. M., Zahir Irani and Weerakkody, V. 2016. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70: 263-286.
- Smit, S., Amsterdam, T. T., Munich, S. L., Washington, D. C. and Manyika, J. 2020. The future of work in Europe. McKinsey Global Institute, Discussion Paper, 52pp.
- Stanly Wilson, S. R. 2019. Twitter data analysis using hadoop ecosystems and apache zeppelin. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(3), 1490-1498. <https://www.researchgate.net/profile/Stanly-Wilson/publication/338198026>
- Suguna, S. and Devi, K. 2016. Improvement of HADOOP Ecosystem and Their Pros and Cons in Big Data. *International Journal of Engineering and Computer Sc.*, **5(5)**: 16680-16685.

- Suryanarayana, V., Sathish, B. S., Ranganayakulu, A. and Ganesan, P. 2019. Novel weather data analysis using Hadoop and MapReduce—a case study. In: *5th International Conference on Advanced Computing & Communication Systems (ICACCS)* (pp. 204-207). IEEE.
- Teixeira, L., Xambre, A. R., Alvelos, H., Filipe, N. and Ramos, A. L. 2015. Selecting an Open-Source Framework: A practical case based on software development for sensory analysis. *Procedia Computer Science*, **64**: 1057 – 1064.
- Thennavan, S. and Thasian, T. 2013. Work Force and Discrimination towards Male to Female Transgender by the Indian Society. *The Indian economic journal*, **2**: 331-338.
- Trajkovska, K. 2022. Software Engineering Statistics: 21 Figures You Should Know About. *Seed Scientific*, 22:7. <https://seedscientific.com/software-engineering-statistics/>
- Trinkenreich, B., Wiese, Igor, W., Sarma, A., Gerosa, M. and Igor, S. 2021. Women's Participation in Open Source Software: A Survey of the Literature. *Computer Science & Software Engineering*,
- Tsai, C.W., Chin-Feng L., Han-Chieh, C. and Athanasios, V. 2015. Big data analytics: A survey. *Journal of Big Data*, **2(1)**: 21.
- Venugopalan, A. and Verma, P. 2020. India Inc opens doors to transgender employees. *Economic Times*, 1(3)
- Verma, C. 2023. Software Engineer Salary in India: Per Month, Fresher's average Salary in India. [Collegesearch.in/articles/software-engineer-salary](https://collegesearch.in/articles/software-engineer-salary). *Software Engineering*, 153/4: 299-314.
- Wolfert, S., Lan G., Verdouw, C. and Marc-Jeroen, B. 2017. Big Data in Smart Farming – A review. *Agricultural Systems*, 153: 69-80.
- Yadav, D. and Chandra, U. 2017. Modern Technologies of Big Data Analytics: a Case study of Hadoop Platform. *Intl. Journal of Emerging Trends and Technology in Computer Science*, **6(4)**: 44-51.
- Yao, D. and Chen, Y. 2020. Design and implementation of log data analysis management system based on Hadoop. *Journal of Information Hiding and Privacy Protection*, **2(2)**:59-65.